ISO/IEC 11179-2 Committee Draft 1

Information technology -- Specification and standardization of data elements

Part 2:
Classification for data elements

Contents

Foreword

ISO (the International Organization for Standardization) and IEC (the International Electrotechnical Commission) form the specialized system for worldwide standardization. National bodies that are members of ISO or IEC participate in the development of International Standards through technical committees established by the respective organization to deal with particular fields of technical activity. ISO and IEC technical committees collaborate in fields of mutual interest. Other international organizations, governmental or non-governmental, in liaison with ISO and IEC, also take part in the work.

In the field of information technology, ISO and IEC have established Joint Technical Committee 1, ISO/IEC JTC1. Draft International Standards adopted by Joint Technical Committee 1 are circulated to national bodies for voting. Publication as an International Standard requires approval by at least 75% of the national bodies casting a vote.

This Part 2 of the International Standard ISO/IEC 11179 was prepared by the Subcommittee ISO/IEC JTC1/SC14, Data Element Principles.

Other parts of the International Standard ISO/IEC 11179 are:

Part 1: Framework for the Generation and Standardization of Data Elements;
Part 3: Basic Attributes of Data Elements;
Part 4: Rules and Guidelines for the Formulation of Data Definitions;
Part 5: Naming and Identification Principles for Data Elements; and
Part 6: Registration of Data Elements.

Introduction

This Part of this International Standard provides the basis for documenting, through a specific set of attributes, the classification aspects of data elements. There are many efforts underway to devise classification schemes and to use the schemes to build and populate classification structures. For the purpose of this standard, the following are all considered types of classification schemes of varying classification power: key words, thesauri, taxonomies, and ontologies. These classification schemes have potentially great utility when associated with various aspects of data elements.

There are several purposes for applying classification to data elements. Classification assists users to find a single data element from among many data elements, facilitates data administration analysis of data elements and, through inheritance, conveys semantic content that is often only incompletely specified by other attributes, such as names and definitions.

The classification schemes accommodated in this Part have utility for
- deriving and formulating abstract and application data elements,
- ensuring appropriate attribute and attribute-value inheritance,
- deriving names from a controlled vocabulary
- disambiguating
- recognizing superordinate, coordinate, and subordinate data element concepts
- recognizing relationships among data element concepts and data elements
- assisting in the development of modularly designed names and definitions.

Each type of classification scheme mentioned above has particular strengths and weaknesses, and provides the foundation upon which particular capabilities can be built. Keywords, for example, are a quick way to provide users some assistance in locating potentially useful data elements. A thesaurus provides a more structured approach, arranging descriptive terms in a structure of broader, narrower, and related classification categories. A taxonomy provides a classification structure that adds the power of inheritance of meaning from generalized taxa to specialized taxa. Ontologies, with associated epistemologies, can provide rich, rigorously defined structures (e.g., directed acyclic graphs with multiple inheritance) that can convey information needed by software, such as intelligent agents and mediators, that are useful in the provision of intelligent information services.

The term *data element* refers to *data element type;* the shorter term is used for convenience.

Information technology—Specification and standardization of data elements
Part 2: Classification for data elements

1 Scope

This Part of the International Standard provides procedures and techniques for associating data with classification schemes. Several components of data elements invite classification—components covered by this International Standard—include object classes, properties, representations, value domains, and data element concepts, as well as data elements themselves. The procedures and techniques specified in this International Standard will enable Registration Authorities to apply classification schemes that better enable one to

- analyze object classes, data element concepts, and data elements,
- make comparisons within the following categories: object classes, properties, representations, data element concepts, and data elements,
- reduce the variety of data element concepts and data elements,
- identify, describe, and define data element concepts and data elements unambiguously,
- assist in the analysis of data elements for the purpose of assigning registration status,
- address synonym and homonym problems,
- retrieve data element concepts and data elements from a data register,
- recognize relationships among data element concepts and data elements

The preparation of this International Standard has also been prompted by the need for standardized data design procedures that will ensure the emergence of data elements capable of supporting electronic data interchange.

This Part of the International Standard develops a set of principles, methods, and procedures for specifying what is needed (at a minimum) to document the association between the various components of a data element and one or more classification schemes. This includes the names, nonintelligent identifiers, definitions, and other aspects of the classification scheme and its contents. These can be captured through use of a set of attributes. Particular attributes are specified in this Part of this International Standard, along with a structure for the contents of these attributes. Users may extend the set of attributes as necessary. Additional information may accompany a taxonomy or ontology; for example, to provide a suggested set of qualifiers that could be applied to the object class, property, or representation taxa to more fully qualify the classification of the particular data element. This Part builds upon and utilizes the basic attributes specified in Part 3 of this International Standard.

An example in this standard shows how selected components of data elements can be associated with a classification scheme through the attributes specified in this International Standard. Use of

one or more classification schemes is intended to provide a sound conceptual basis for the development of data having enhanced semantic purity and design integrity.

This International Standard does not establish a particular classification scheme as preeminent. Sanction of a particular taxonomic approach and/or a particular epistemology is also beyond the scope of this International Standard. These are addressed by other standards committees and/or tend to be tailored to a particular domain of discourse. The power of the classification scheme and the utility of the content are appropriate areas for competition. Other standards committees are developing or have developed normative languages for use in classification and/or particular techniques and structures that can be accommodated by this International Standard. For example, the National Information Standards Organization (NISO) has developed a standard for development of a thesaurus. It is appropriate for each classification structure to be documented as to how it was developed and how it can be extended and maintained. Such attributes could be added, by the principle of extensibility, to the attributes specified in this International Standard. They are not, however, included here.

Each Registration Authority, as described and specified in Part 6 of this International Standard, may classify components of a data element according to the classification schemes, structures, and content that it deems appropriate. In documenting the classification aspects of data elements, the Registration Authority shall do so according to the principles, methods, procedures, and attributes specified in this International Standard.

2 Normative References

The following standards contain provisions that, through reference in this text, constitute provisions of this Part of the International Standard. At the time of publication, the editions indicated were valid. All standards are subject to revision, and parties to agreements based on this Part of the International Standard are encouraged to apply the most recent editions of the standard indicated below. Members of IEC and ISO maintain registers of currently valid International Standards.

None.

3   Definitions

For the purposes of this Part of the International Standard, the following definitions apply.

3.1 administered component: A component for which administrative attributes are collected. See ISO/IEC 11179, Part 3, Clause 5.2.4 for a list of administrative attributes.

3.2 attribute: A characteristic of an object or entity.

3.3 classification scheme: An arrangement or division of objects into groups based on characteristics which the objects have in common, e.g. origin, composition, structure, application, function, etc.

3.4 classification scheme item: The components of content in a classification scheme. These may be the nodes of a taxonomy/ontology, the terms of a thesaurus, etc.

3.5 classified component: Any administrative component of a data element that can be classified in one or more classification schemes. These components include the object class, property, representation class, data element concept, value domain, and data element.

3.6 data: A representation of facts, concepts, or instructions in a formalized manner suitable for communication, interpretation, or processing by humans or by automatic means (ISO 2382-4).

3.7 data element: A unit of data for which the definition, identification, representation, and permissible values are specified by means of a set of attributes.

3.8 data element concept: A concept, which can be represented in the form of a data element, described independently of any particular representation.

3.9 name: The primary means of identification of objects and concepts for humans.

4 Classification Attributes for Data Elements

Attributes shall be utilized to associate various classification schemes with selected components of data elements. These components include object classes, properties, representations, and data element concepts, as well as data elements themselves. As described in Part 1 of this International Standard, data element concepts are comprised of a relationship between one or more object classes and a property. A data element is formed when a form of representation (e.g., a numeric code) is chosen for the data element concept. Each of the components may be associated with classification schemes as described below.

4.1 Keywords

Keywords are included as basic attributes in Part 3 of this International Standard. They can be applied to object classes, properties, representations, data elements and to data element concepts. Part 3 describes keywords as follows:

> name   :   **Keyword**

definition : One or more significant words used for retrieval of data elements.
obligation : Optional
data type : Character string
comment : This attribute can be used for recording keywords (search keys) associated with the data element in question.

For the purpose of this International Standard, "controlled word lists"—wherein each word in a list of words is limited (controlled) to a particular meaning—can be recorded in the same manner as keywords. While key words are very weak as a classification scheme, they are nonetheless very useful and can be recorded as a "classification scheme type" for use within this standard.

4.2 Thesaurus terms

Thesaurus terms can be associated with data elements and data element concepts. The structure of the thesaurus is not specified in this International Standard. The attributes specified in clause 4.4, below shall be utilized for capturing thesaurus information related to data elements or data element concepts.

4.3 Taxonomy and ontology taxa

The taxa in taxonomies and ontologies may be related to the following classified data registration components: object class, property, representation class, and data element concept. The attributes listed in clause 4.4 below shall be utilized to capture this information.

4.4 Classification attributes

The following attributes may be used for recording the classification information about any component of a data element:
Classified component
Classification scheme type
Classification scheme name
Classification scheme version
Classification scheme item type
Classification scheme item value

For example:
Classified component = Object Class (e.g., fruit fly)
Classification scheme type = Taxonomy
Classification scheme name = International Taxonomic Information System (Biological Taxonomy)
Classification scheme version = 2.1
Classification sheme item type = taxon identifier
Classification scheme item value = 5411

or
Classification scheme item type = taxon term
Classification scheme item value = Drosophila

The relationships among the classification attributes are depicted in Figure 4.1. Additional attributes may be utilized where necessary.
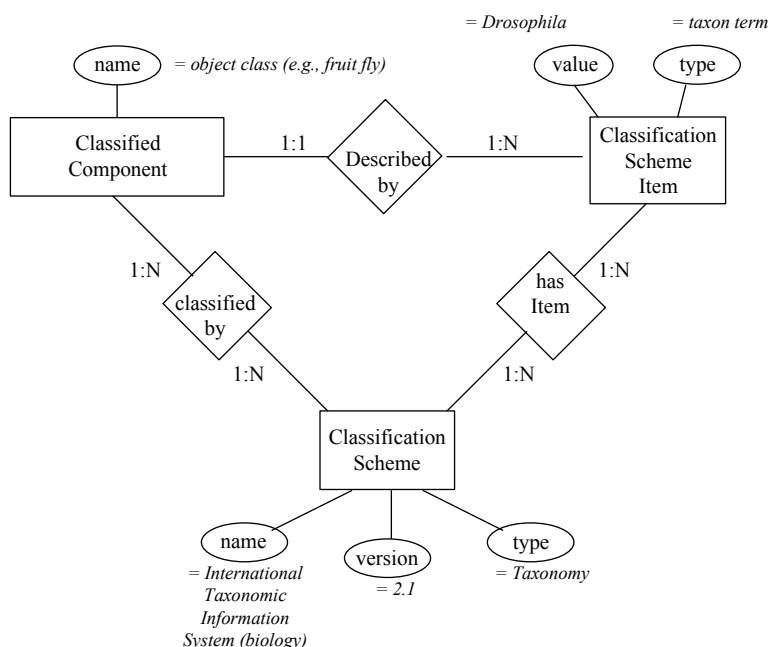

Figure 4-1. Classification Attributes for Administered Components, and an Example of their Usage

4.5 Classification attribute descriptions

4.5.1 name   :   **Classified component name**
       definition :   The name of the data element component that is subject to classification, such as object class, property, representation class, value domain, data element concept, and data element.
       obligation :   Conditional
       condition :   Required if the classification items are to be recorded.
       data type :   Character string
       comment   :   Example: Object Class

4.5.2 name   :   **Classification scheme type**
       definition   :   The type of classification scheme from which classification items are drawn.
       obligation   :   Conditional
       condition   :   Required if the classification items are to be recorded.
       data type   :   Character string
       comment   :   Example: Taxonomy

4.5.3 name   :   **Classification scheme name**
       definition   :   The name of the particular classification scheme from which the classification
                        items are drawn.
       obligation   :   Conditional
       condition   :   Required if the classification items are to be recorded.
       data type   :   Character string
       comment   :   Example: Kenworthey's Taxonomy

4.5.4 name   :   **Classification scheme version**
       definition   :   The version of the particular classification scheme from which the classification
                        items are drawn.
       obligation   :   Conditional
       condition   :   Required if the classification items are to be recorded.
       data type   :   Character string
       comment   :   Example: " Version 2.1"

4.5.5 name   :   **Classification scheme item type**
       definition   :   The type of the "classification scheme item value" in 4.5.6, below.
       obligation   :   Conditional
       condition   :   Required if the classification items are to be recorded.
       data type   :   Character string
       comment   :   Examples: "Unintelligent identifier", "Term" or "Definition"

4.5.6 name   :   **Classification scheme item value**
       definition   :   An instance of the name, or identifier, or definition, etc. of a classification scheme
                        item.
       obligation   :   Optional
       data type   :   Character string
       comment   :   Examples: "5411" is an example of a value of type "unintelligent identifier" for a
                        taxon within a taxonomy. "Drosophila" is an example of a value of type "term"
                        used to identify a node in a taxonomy.